

BIROn - Birkbeck Institutional Research Online

Nunan, Daniel and Knox, S. (2011) Can search engine advertising help access rare samples? *International Journal of Market Research* 53 (4), pp. 523-540. ISSN 1470-7853.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/17838/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

Can search engine advertising help access rare samples?

Authors:

Dr. Dan Nunan & Professor Simon Knox

Cranfield School of Management

Abstract

In last decade, there has been an explosion in the use of online survey tools. Online data collection tools have lowered the cost of data collection and removed barriers to entry for carrying out research. Whilst a number of questions have been raised about the general reliability of internet survey research, one specific use of the web for survey work has been in reaching niche populations that are difficult to access using traditional survey tools, so called 'rare samples'. In this paper, we present an approach to accessing such hard-to-reach populations using search engine pay-per-click (PPC) advertising. We carried out a study that makes use of pay-per-click (PPC) advertising on search engines as an alternative means of developing sample for a hard-to-reach group of health consumers. Based on a sample of 466 consumer responses, we discuss the effectiveness of this technique for reaching such rare populations.

Introduction

Technological changes through the ‘noughties’ have enabled a huge increase in the use of web-based surveys by researchers. Trends over the last decade have been driven by the adoption of the web and mobile devices as dominant forms of communication and information distribution. In many countries, web-access can now be considered as ubiquitous as fixed-line telephony once was (Geoghegan, 2009). At the same time, technology has also challenged existing modes of data collection. For example, increasing use of caller id devices has substantially increased non-response rates for telephone research (Jarvis, 2002). Despite the popularity of online surveys there have also been increased concerns over survey error, particularly around selection and non-response bias (Sharot, 2010). Yet, online surveys have also been seen as an effective means of reaching empirically underrepresented and hard-to-reach populations (Koch & Emrey, 2001). The last few decades have seen a reduction in demand for general samples of consumers and an increase in research using highly specific groups, or ‘rare populations’ that represent a small proportion of an overall population (Sudman & Blair, 1999). Reaching such populations through traditional random sampling techniques may require significantly more resources than are available to build a sufficient sample size. In these circumstances where samples would otherwise be unavailable, pragmatism forces the use of more purposive, and perhaps less desirable, sampling techniques.

In this paper, we review the current challenges in gaining responses from empirically underrepresented samples. We consider the opportunities that are presented both by changing technology and shifts in consumer information-seeking behaviour to present a more effective approach for gathering such hard-to-reach samples. Using a pay-per-click (PPC) search engine technique as a data collection tool across a sample of 466 consumers, we analyse the effectiveness of this technique for accessing niche samples. We further consider the effectiveness of this approach in terms of i) the use of incentives ii) invitation wording iii) response cost and iv) its appropriateness for gathering samples. We then consider in broader terms the applications of this approach, together with any related limitations.

The Growth in Web-Based Data Collection

The growth in the use of the web for data collection can be seen as a function of both the enabling capabilities of web-based research tools and the limitations of more traditional forms of research. In the context of declining response rates to traditional forms of surveys, web-based surveys provide an alternative. Public response rates for surveys have been on a downward trend since the 1950s (Steeh, 1981) and the rate of decline has been increasing since the mid 1990s (Curtin et al., 2005). In their paper Random Digit Dialing based telephone surveys have replaced mail surveys *non-response due to rejection* with *non-response due to non-contact*. Whilst many research professionals have sought to identify aspects of methods that are the source of non-response, Curtin et al. (2005: 97) note that there may be a broader underlying problem:

“The most promising explanation to account for the ... increase in both refusals and noncontacts over the past 25 years may be the rapid growth in sales and survey phone calls during the period.”

In summary, respondents are unable to differentiate between invitations for ‘professional’ market research and those for sales and survey calls. In turn, consumer willingness to receive cold-calls has declined. National ‘do-not-call’ lists have been introduced in many countries including the US, Australia and Canada. These schemes are designed to prevent unwanted telemarketing calls and have specific exemptions for market researchers. Despite initial fears of harm to the market research industry, research has found that respondents are now less likely to hang-up the phone to researchers (Yost et al., 2005). However, in the US the system has been subject to telemarketers and political campaigners using the cover of research to include partisan messages in ‘push-polls’ (Farbman, 2005).

Kellner (2004) highlights a number of reasons why online surveys may be preferred in practical terms over telephone approaches:

- 1) Lifestyle choices are resulting in reduced contact rates as fewer people are home at the sort of times when market researchers call. For those that are at home, there is increasing concern over demographic differences between those that are out. For example, people who are more likely to be at home may also be more likely to be unemployed.

- 2) Technology, through caller-ID, voicemail and increased use of mobile telephony replacing fixed line phones has resulted in greater difficulty in reaching potential respondents. Some researchers have gone as far as to question the long-term future of telephone research given these trends (Curtin et al. 2005).
- 3) The existence of 'a spiral of silence' whereby respondents 'consciously or subconsciously conceal responses that they fear are 'socially unacceptable' (ibid: 8). An example of this in telephone polling has been a consistent bias towards some political parties and away from others.

The question of whether issues of non-response are specific to online surveys or reflect wider sociological issues in attitudes to survey responses, is an area that requires further research (Evans and Mathur, 2005). However, the sociology of non-response suggests that the issue is not that technology is preventing researchers from reaching respondents; rather it is a number of wider changes in society that are causing respondents to avoid participating in research (Tourangeau, 2006). People feel too busy to participate, even though they have more free time than before (Robinson and Godbey, 1997); the growth in multitasking has led them to think that they are busier. It appears that they are less interested in helping others, for example, by completing surveys (Tourangeau, 2006). Finally, people are erecting barriers against unwanted intrusions. Reactions to junk mail and telemarketing are one aspect of this, as are more physical manifestations such as the rise in gated communities (Tourangeau, 2006).

This combination of socio-cultural and technological factors has led to a cycle of non-response that has substantially increased the cost of developing high quality samples (Sudman & Blair, 1999). Faced with such pressure on costs it is not surprising that researchers have turned increasingly to the use of online survey tools. One such tool, the US-based 'SurveyMonkey', is now collecting over 300 million individual responses to surveys hosted on its systems each year (Survey Monkey, 2009).

Challenges with online data collection

Survey methods that rely on stopping respondents in the street, sending mail surveys or phoning people at home are from an era where these were the dominant forms of communication. The US postal service still ships around 14 billion items a month, although this volume has been in steady, long-term decline. This can be compared to 88 billion searches on the Google search engine each month (Comscore, 2009); access to the Internet in the home is now approaching landline penetration in the US and UK and exceeds it in many countries (Geoghegan, 2009). The average Internet user in North America and Europe makes around 110 searches a month, a number that increases to 130 a month for Latin America (Comscore, 2009). One significant weakness of web-based research is that it often requires the invitation to complete research to be distributed via email. Whilst some early guidelines on carrying out survey research suggests the use of multi-modal techniques, such as sending invitations by email and post (Dillman, 1991), it should not be forgotten that in the digital world email and the web are different modes. An example of such a challenge is the growth in smartphones which is resulting in email increasingly being checked on a mobile device with no easy way of clicking-through to a survey designed for display on a computer screen.

Information theory suggests that we can divide the way that consumers use the web into two categories. Firstly the undirected browsing of websites, for example by reading a newspaper is one sort and a second type is a goal-directed information search, whereby consumers are surfing the web with a specific goal in mind. Goal-directed searches typically involve some form of use of a search engine to help guide and direct the search (Martzoukou, 2005). From the perspective of marketing professionals, search engines present two opportunities for reaching customers. The results generated from a consumer search query into a search engine, such as Google or Bing, can be divided into organic and paid results. Organic results are the results generated by the search engine algorithm based on its analysis of the relevance of the site to the search query. These are not paid for advertising and are the results that appear down the left-hand side of a page. Many firms are now allocating a significant proportion of their marketing budgets to search engine optimization (SEO) where they manage and alter their online presence to fit the requirements of search engines. Getting a listing at the top of a major search engine (Google, Bing or Yahoo) is now seen as a form of validation of a site by the search engine (Thurow, 2006); if a site appears at the top of Google results then it is perceived as having

some underlying quality. It has been suggested that focusing resources on achieving better organic search engine rankings is the best investment that firms can make in achieving brand impact online (Noaman, 2006). This can create some incongruence in that consumer expectations of brand validation through search engine placing may be misplaced. For example, a smaller, lower standard hotel may be able to invest a large amount of resources to get to the top of a search engine ranking for a 'luxury hotel in location x' search above more well-known premises (Dou et al. 2010).

We argue that the search engine is the *de facto* 'high street' of the web. Advertisers focus on search engines as the starting point for purchase intent. The majority of visitors to web sites arrive there via a search engine, rather than by typing in a specific URL or by using a bookmark or another form of link (Telang et al., 2004) which seems to highlight the growing importance of search engine optimization (SEO) to marketing professionals. We argue that looking at behaviour in the search box provides a window through which to view what individuals are looking for online. For research focusing on a specific product or service category, this also provides an opportunity for access to a sample that is closer to the ideal of a random sample than is currently available through sampling techniques based on convenience. A decade ago Taylor (2000) highlighted that the internet would require fundamental changes in the way that survey data are collected and analysed. Top of the list of these changes is that internet research is not based on probability sampling but on volunteer sampling. We believe that alternatives to email as a survey invitation mechanism are required given the inherent, and increasing, limitations in the email platform. In the following section we outline and investigate an alternative technique.

Using the web to access rare populations

The challenges in researching hard-to-reach or 'rare' populations are long-established. Techniques such as random digit dialing can be inefficient requiring a large volume of initial screening thus creating a high cost-per-response (Ghosh-Dastidar et al. 2009). One study into rare populations found that a sample of 30,720 would be required to achieve 95% confidence on a relevant sub-sample of just 768 people (Mathy et al. 2002).

Rare populations can be characterised as lacking adequate sample frames, consequently,

researchers are faced with using non-probability sampling techniques. Examples of non-probability research using hard-to-reach populations includes minority religious groups (Reed, 1975; Jelen, 1992; Wald, 1988), war veterans (Rothbart, et al, 1982) and gang members (Valdez and Kaplan, 1999). In some areas, such as health services, the combination of complex ethical requirements and the difficulties in effectively operationalising research into rare samples has resulted in academic researchers investigating other service contexts (Brown, Crawford & Carter, 2006). Winship & Mare (1992:327) summarise the essential paradox with such populations:

“non-random selection is both a source of bias in empirical research and a fundamental aspect of many social processes.”

The factors that make certain populations worth studying also make them difficult and expensive to access using ‘textbook’ research methods. The rarer the population, the greater the potential cost and difficulty in building a sample.

For these reasons, researchers requiring rare samples are using the web to reach a greater concentration of eligible respondents due to the ability to make use of filtering variables not available in random dialing or mailing approaches (Ghosh-Dastidar, 2009). Whilst such techniques may be effective in terms of making available *some* responses from a sample population, the nature of these convenience samples make them particularly susceptible to the problems of non-response, particularly when survey invitations are done via email. We now consider an alternative approach that makes use of search engine traffic to deliver highly targeted traffic for the accessing of rare samples.

A new technique for sampling online surveys

In this study we make use of the Pay Per Click (PPC) advertising system to access an otherwise hard-to-reach sample of health service consumers. PPC business models are the revenue generator for many leading search engines and were pioneered by the market leader Google. Unlike traditional advertising models for which advertising is charged based on the number of people viewing an advert, PPC models are based only when a viewer takes action. In this case the action is clicking on an advert. PPC models enable more effective allocation of resources by advertisers; their increasing popularity amongst advertisers stems from the apocryphal marketing problem of “50% of our advertising

budget is wasted, I just don't know which 50%". PPC on its own is not a research tool but combined with the targeting available through use of messaging variables, it opens up capabilities for extremely targeted approaches to sampling. Figure 1 highlights these variables and their functions.

| Variable | Functions of Variable |
|-----------------|--|
| Geography | Limit message to specific countries or for larger countries, specific regions. |
| Search Terms | A range of Boolean search options enable the targeting of messages at specific search terms, or combinations of terms. |
| Time | Messages can be shown or excluded during certain times of days. |
| Budget | The order in which PPC messages are shown is dependent on the budget allocated to each message. A proprietary auction system is used to decide the order in which messages are shown |

Figure 1. PPC variables and their functions

Messages are limited to a specific format which is a headline, two lines of text and a URL with a maximum of 95 characters being allowed (excluding the URL). Once the variables have been selected and the message written, the result is an advert that appears either above or to the right hand side of normal search results as is shown in figure 2. The position of the message on the screen is determined by the budget set by the advertiser and the relevance of the message to the audience. In other words, it is a combination of the amount that an advertiser is willing to pay for a click on their message and the likelihood of someone finding the message compelling enough to click on. The precise positioning cannot be guaranteed as it is calculated through an auction system with the underlying formulae a closely guarded secret by the search engine companies.

market research Search SafeSearch moderate ▼

About 169,000,000 results (0.19 seconds) Advanced search

Market Research Company
www.BrandInstitute.com Effective Market Research Solutions For All Types Of Businesses.

Market Research
www.tlmc.co.uk Competitive B2B telephone market research services. Contact us now!

Quality Market Research
www.Adroit-e.co.uk High Tech, Quants Market Research Accurate Data, Rapid Results

Market research - Wikipedia, the free encyclopedia ☆
Market research is any organized effort to gather information about markets or customers. It is a very important component of business strategy. ...
History - Market research for business ... - Financial performance
en.wikipedia.org/wiki/Market_research - Cached - Similar

Marketing research - Wikipedia, the free encyclopedia ☆
Marketing research is the systematic gathering, recording, and analysis of data about issues relating to marketing products and services. ...
en.wikipedia.org/wiki/Marketing_research - Cached - Similar

Market Research Society ☆
With members in more than 70 countries, MRS is the world's largest professional association representing providers and users of market research and business ...
www.mrs.org.uk/ - Cached - Similar

Market Research Reports - Business Market Research Reports ... ☆
Business Market Research Reports and Industry Analysis available from Market Research.
www.marketresearch.com/ - 34 minutes ago - Cached - Similar

Market Research, By DJS Research Ltd ☆
Written By: DJS Research Ltd, full service market research agency. Market Research is a systematic, objective collection and analysis of data about a ...
www.marketresearchworld.net/index.php?Itemid - Cached - Similar

Market Research Company
Qualitative & quantitative market research service
www.djsresearch.com

Market Research
Access The Latest Market Research Reports - Immediate Download
www.ReportLinker.com

Market Research Companies
Compare Quotes from 4 leading UK Market Research Companies.
ApprovedMarketResearchComp.co.uk

NEMS Market Research
Quality & cost effective surveys Telephone, in-home & in-street
www.nemsmr.co.uk

Paid Online Surveys
Convert points to great vouchers for fun and innovative surveys
www.justtheanswer.co.uk

Your time is precious
That is why we always reward you for taking part in our paid surveys
www.ValuedOpinions.co.uk

Market Research Reports
Market Sizes and Trends analysed.

Figure 2. Example of message placement.

We believe that the availability of PPC technology, in conjunction with the ability to display targeted messages in certain search engines, provides opportunities for building research samples for rare populations. It allows the researcher to leverage context specificity in that search terms tell you something about the intent of the searcher. For example, if you seek to carry out research only on a population of cat owners, you could place a message only in front of people searching for the word ‘cat’. You could also narrow this down to cat owners in the United Kingdom, but exclude people searching for ‘CAT cable’ (a type of computer cable).

Experimental Study

The context for this study was a wider research project on consumer engagement with private sector health services in the UK. This project was related to investigating consumer understanding of risks across a range of specific health products. The context was non-clinical (i.e. this wasn’t about investigating people who were ill) and, as a result, respondents were sought from a general population rather than a specific medical sample. The sample required both a geographic filter for the UK and to limit responses to consumers who had knowledge of certain specific health services. Health services are a

difficult area to research due to the complexities of selecting respondents, achieving ethical approval and the prospect of a high non-response for unsolicited face-to-face invitations related to health matters.

To meet the sample requirements for this study, we investigated the use of the PPC sampling technique as part of an experiment with the goal of exploring the effectiveness of this approach. Specifically, we sought to investigate the following areas:

1. The impact of message wording on response rate.
2. The impact of incentives on response rate.
3. The cost effectiveness of this approach versus other web research methods.
4. Overall response rates versus other web research methods.

In this study, we limited our research to the use of Google as the dominant search engine in the UK, although the principles are applicable to other major search engines. Indeed, it has even been suggested that because of these difficulties in gaining access to samples in professional healthcare, academic researchers have tended to focus on professional groups that are easier to access, such as lawyers (Brown et al., 2006).

The first step was developing a comprehensive list of search terms which could then act as a gateway for particular health screening messages. Search terms were refined through a three stage process. A panel of health experts was asked to come up with phrases and terms that could be used to contextualize our research. This was tightly focused on words that matched the specific types of health screening services that were the subject of this study. Once this list had been developed, it was entered into a keyword analysis tool provided by Google that applies an algorithm to suggest missing terms based on overall search patterns. These additional words and phrases were broader matches less directly related to the research context but still relevant in terms of the audience. An example of such a phrase was “private doctor”.

In coming up with a list of such words, the aim was to arrive at the broadest range of relevant terms possible whilst maintaining the efficiency of the PPC technique by ensuring that search phrases were not seen by those who were non-members of the target sample group. On completion of the study we cross-checked responses to ensure that there was broad similarity in response rate between keywords. We also included a check

question within the study to verify that respondents from these search keywords were legitimate members of the sample population.

Next, we created four messages within four separate categories to test the effect that message wording had on response rates (figure 3). Given the restrictions over message size, messages were written in the manner of a generic survey invitation without detailed information on the purpose of the survey. A feature within the Google PPC system allows for the initial testing of the messages on a small sample to determine which was the most effective in terms of response rate. The most successful message was then automatically used in the main study. Each message led through to an explanation of the purpose of the study and information on a small incentive - the opportunity to win a £50 Amazon.co.uk voucher. Whilst the use of incentives has been found to increase response rates in online surveys (Cobanoglu & Cobanoglu, 2003), there are also questions of potential bias when using incentives in online surveys. Incentives can result in multiple completions and offering unfair incentives can alter attitudes and responses to survey questions (Larson & Sachau, 2009). In this study, we took a number of technical steps to help limit responses to a single response per individual. Whilst there is always the risk of some workaround may enable multiple completions, we believe that precautions taken were sufficient in this case. Additionally, the PPC approach helps to deter more opportunistic completions by ‘professional survey takers’ as the survey invitation is only discoverable by those searching on specific keywords. Furthermore, in this study we deliberately used a modest incentive that did not guarantee financial gain on completion.

| Message Category | Message Wording |
|--|--|
| Financial incentive + Research Credentials | Health Screening Research Key Academic Survey. Take part and win an amazon.co.uk voucher. www.[anonymised for review] |
| Information incentive | Health Screening Research Find out more and take part in this university research project. www.[anonymised for review] |
| Research credentials | Health Screening Research Key Academic Survey on health |

| | |
|---------------------|---|
| | tests needs your input. www.[anonymised for review] |
| Financial incentive | Health Screening Research Take a survey for a chance to win an amazon.co.uk voucher. www.[anonymised for review] |

Figure 3. Message Wordings

Findings

The study was run over a four-day period during which the message appeared 141,058 times in search results. The full survey invitation was viewed 1,663 times and resulted in 444 survey responses. The question arises of the definition of response rates in this context. What form of non-response is analogous to an ignored message in search results? If we draw a direct link between a message being shown in search results and response rates they appear very low (0.3%), yet for those who read the full survey, completions were 27%. We argue that a non-response to a message in search results is not the same as a non-response to a mailed survey or a phone call that is not picked up. A typical Google search page can include links to up to 32 sites, up to 11 of which will be some form of PPC result. We argue that the response rate is a function of the level of attention that an individual message receives. Broder (2002) notes that only around half of web searches are for informational reasons, the rest being transactional or navigational. This is important as the willingness to respond to an informational message, such as a survey invitation, is likely to be limited where the purpose of the search is not informational. Individual response rates are also highly influenced by the number of messages on each page and the positioning of the message. Eye tracking studies, which measure the position of the eye relative to the screen, have attempted answer the question of what percentage of respondents actually saw the messages. One study on attention paid to Google search (Eyetoools, 2005) found that the percentage of an audience who viewed varied from 10% through to 80%-100% depending on the position of the message (figure 4).

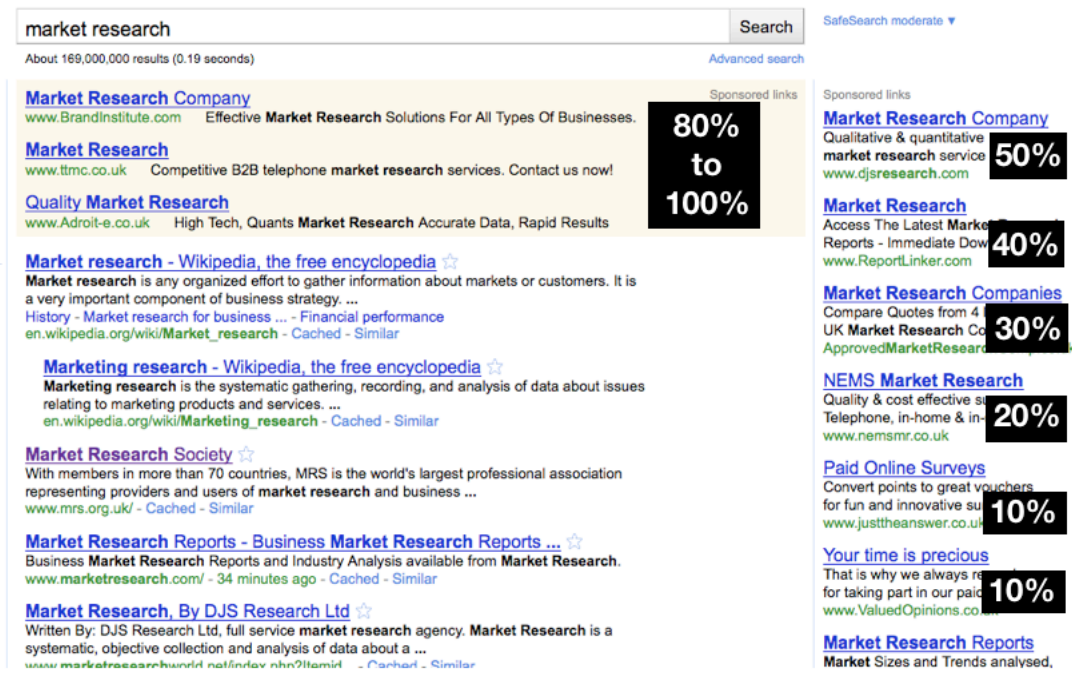


Figure 4. % of audience who view each PPC message depending on position

Whilst the health context of this study could be considered a rare sample, even this case resulted in a relatively large number of searches in a small period of time. For rarer samples, we believe the response rate is likely to increase due to a smaller number of PPC sponsored links, and greater attention on an individual advert. It is also likely that a greater response rate could have been achieved by further testing and improvement of the message itself so that it was less a generic survey invitation, and more an effective ‘call to action’ that could compete for attention against other PPC messages and page content.

The response rates for specific messages are shown in figure 5. Note that these response rates were calculated as part of the pilot test, and only the first message was used in the full survey. The findings suggest that in this case, the financial incentive alone produces a relatively small motivation to respond. Whereas offering information and emphasizing that the study was for academic purposes increased this response rate. This may be due to the nature of the incentive as previous research indicates that promised incentives are far less likely to influence respondents than prepaid incentives, such as including a \$5 payment in a survey mail-out (Berk et al., 1987). It also suggests that someone carrying

out a search on a search engine has more of an information-seeking goal and is, therefore, in the correct ‘mode’, whereas a person being invited via an email invitation needs to shift modes and that a financial incentive may be more effective in this task. In order to demonstrate criterion validity, a set of demographic data was collected from each respondent. This included age group, gender and education level. Each variable was cross-tabulated and checked against data on overall internet use and existing profile data on consumers of the health services featured in this research.

| Message Category | Response Rate |
|--|----------------------|
| Financial incentive + research credentials | 1.18% |
| Information incentive | 0.42% |
| Research credentials | 0.61% |
| Financial incentive | 0.16% |

Figure 5. Message Response Rates

The different ‘costs per click’ for each message are displayed in figure 6 below. As with figure 4 the first message represents the cost per click for the study as a whole whereas the other message costs were determined from the test.

| Message Category | Average Cost per Click |
|--|-------------------------------|
| Financial incentive + Research Credentials | £1.33 |
| Information incentive | £0.82 |
| Research credentials | £1.82 |
| Financial incentive | £0.74 |

Figure 6. Average Cost per Click

The average cost was a function of the relevance of the message to the search and the competition from other advertisers for space against that search. Due to the proprietary nature of the algorithm, it was not possible to carry out further analysis on these numbers, although we noted that the ‘research credentials’ message resulted in a number of advertisements for university courses next to it which could explain the higher cost. Overall, this approach delivered a cost per completed response of £4.98 which was an acceptable cost, given the nature of the research sample.

Discussion

We believe this approach has a number of potential advantages for gaining access to hard-to-reach samples, in particular those convenience samples generated from email invitations. Firstly, given the scope of coverage of search engines, it provides a practical means of reaching very highly targeted sample populations at a low overall cost in a fast time. Due to the very large number of searches{over 110 *billion* per month on the three leading search engines alone (comScore, 2009)}, this technique provides a means of accessing a huge range of people, including many who may not be readily accessible by any other means. For example, in the case of health service research the internet provides a perception of anonymity that is less present in other modes of communication. Even in this instance with a geographically targeted sample group, we were able to achieve a large number of responses within a short timeframe. Whilst this rate will vary depending on the specific search terms being used, we believe strongly that PPC does ease access and lowers the barrier to entry for many complex or difficult-to-target research topics. Secondly, this approach negates the need for a multi-modal approach to survey invitations where email is relied upon as a mechanism for inviting people to complete web-based surveys. Harnessing information-seeking intentions through the use of a search engine does seem to reduce the friction felt by respondents in completing the survey and helps reduce the drop-out rate.

Given the novelty of this approach and its use of emerging technology we explore a number of potential issues that had to be consider. The first question is the type of research topic that lend itself to generating a sample via search engines. The research

study into health consumers related here was very much oriented around information seeking and could be readily mapped to search terms. However, as usage of the Internet is by no means universal and search engines do not provide a record of all online behaviour, there are many 'offline' areas. Yet, for those researchers who have decided to carry out web-based surveys, PPC techniques give access not only to a highly targeted sample but also a sample that is already using a web browser and thus in a position to easily respond.

The second question is whether this method improves upon existing non-probabilistic approaches to rare samples. Simple random sampling requires a sample frame where we know something, directly or indirectly, about every respondent (Dorofeev & Grant, 2006). In this respect this technique suffers from the same limitations identified in other forms of web-based survey. However, we believe this technique should be considered by "the empirical pragmatist", a group "who will accept a research methodology (for example, quota sampling) if it consistently produces credible data and which can be shown to 'work'" (Taylor, 2000: 52). The underlying socio-cultural attitudes towards the web and data privacy, combined with the limitations identified in this paper, make the broad availability of high-quality web-based samples very difficult. However, we believe that the use of sampling techniques identified in this paper provides a pragmatic solution to help improve the accessibility of rare samples for web-based surveys. We also believe that with further development, the PPC approach can be improved using augmented sampling techniques, such as snowballing.

There are also a number of technical issues that should be considered. Firstly, this technique requires a systematic approach to the development of lists of search terms. We suggest making use of automated tools to ensure that the list is as comprehensive as possible. Secondly, whilst technology can be used to prevent multiple completions with the specific techniques used, there is currently no way of ensuring that multiple invitations are not received by respondees searching multiple times. Thirdly, there is variation in search engines between countries. In major Internet markets such as Japan and China, Google is not the dominant search engine. This means that researchers may need to make use of a range of search engines in each market to ensure they reach the maximum population, thus increasing the complexity of the sampling approach for multi-

market surveys. Fourthly, PPC approaches rely on bidding against commercial advertisers. Whilst the informational nature of a piece of legitimate market research provides some protection in terms of keeping the cost-per-click lower than for commercial operators, there are likely to be some categories whereby this technique may simply be too expensive. For example, in the US putting messages alongside the term ‘mesothelioma’ could result in cost-per-click of nearly \$100 due to the high volume of lawyers advertising for clients for asbestos litigation (AdGooroo, 2009). Finally, there is the limitation of the short message format that requires the researcher to be able to explain and argue for the purpose of their research in a relatively small number of characters.

Conclusions

In this paper we have demonstrated a technique for accessing rare samples using PPC advertising. As technologies develop, they provide new opportunities for harnessing information. We believe that the approaches outlined in this paper illustrate how one such opportunity can be used effectively in online research. Whilst we have been able to gain a large number of responses quickly in this research, we have identified a number of areas in which overall response rates could be improved in future research.

Whilst some limitations have been discussed with regards to this approach we also believe this paper highlights future opportunities for sampling based on search engines. As with the growth of the internet, the technologies described in this paper are themselves changing, evolving and adapting. The approaches used in this research based on keyword matching may be considered relatively primitive when compared to the behavioral targeting approaches now being developed by the technologists behind the leading search engines. We believe these and other future developments will offer present substantial opportunities for further research into the use of search engines based techniques for sampling. Although the scope of this paper has been focused on rare samples we foresee that as technology evolves opportunities may become available for carrying out effective probabilistic sampling techniques using search engines. In publishing this paper we hope to raise awareness of important, and increasing, role of search engines in consumer information seeking and highlight the opportunities in adapting these technologies for building research samples.

References

- AdGooroo. (2009) Q409 Search Advertising Report: Thinning the Herd.
(http://succeed.adgooroo.com/Q409_Search_Advertising_Report.html).
- Berk, M., Mathiowetz, N., Ward, E & White, A. (1987) The Effect of Prepaid and Promised Incentives: Results of a Controlled Experiment. *Journal of Official Statistics*, **3**, 4, pp. 449-457.
- Biemer, P. & Lyberg, L. (2003) *Introduction to Survey Quality*. Hoboken, NJ: John Wiley & Sons.
- Brown, B., Crawford, P. & Carter, R. (2006) *Evidence Based Health Communication*. Maidenhead: Open University Press.
- Cialdini, R. (1994). *Influence: The Psychology of Persuasion*, NY: Quill
- Comscore. (2009) Global Search Market Draws More than 100 Billion Searches per Month.
(http://www.comscore.com/Press_Events/Press_Releases/2009/8/Global_Search_Market_Draws_More_than_100_Billion_Searches_per_Month)
- Cobanoglu, C. & Cobanoglu, N. (2003). The effect of incentives in web surveys: applications and ethical considerations. *International Journal of Market Research*. 45:4, pp. 475-488
- Curtin, R., Presser, S. & Singer, E. (2005) Changes in Telephone Survey Nonresponse over the Past Quarter Century. *Public Opinion Quarterly*, **69**, pp.87-98.
- de Leeuw, E., & de Heer, W. (2002) Trends in Household Survey Nonresponse: A Longi-

tudinal and International Perspective. In Groves, R., Dillman, D. Eltinge, J. & Little, R. (eds) *Survey Nonresponse*, pp. 41-54. New York: Wiley.

Dillman, D. (1991) The design and administration of mail surveys. *Annual Review of Sociology*, **17**, pp. 225–249.

Dorofeev, S. & Grant, P. (2006). *Statistics For Real-Life Sample Surveys: Non-Simple-Random Samples and Weighted Data*. Cambridge: Cambridge University Press

Dou, W., Lim, K., Su, C., Zhou, N. & Cui, N. (2010) Brand Positioning Using Search Engine Marketing. *MIS Quarterly*, **34**, 2, pp.261-279

Evans, J. & Mathur, A. (2005). The value of online surveys. *Internet Research*, **15**, 2, pp.195-291.

Eyetoools (2005). Enquiro/Eyetoools Google Report.
(http://eyetoools.com/inpage/eyetracking_research.html)

Farbman, R. (2005) Has the Do Not Call Registry Turned Out To Be a Survey Researcher's Best Friend? Edison Research Blog
(http://www.edisonresearch.com/home/archives/2005/07/has_the_do_not.php)

Geoghegan, T. (2009) The slow death of landline phones, BBC News Online.
(<http://news.bbc.co.uk/1/hi/8355504.stm>).

Ghosh-Dastidar, B., Elliott, M., Haviland, A & Karoly, L. (2009). Composite Estimates from Incomplete and Complete Frames for Minimum-Mse Estimation in a Rare Population. *Public Opinion Quarterly*, **73**, pp. 761-784.

Groves, R., Singer, E. & Coming, A. (2000) Leverage-Salience Theory of Survey Participation: Description and an Illustration. *Public Opinion Quarterly*, **64**, pp. 288-308.

Jarvis, S. (2002) CMOR finds survey refusal rate still rising. *Marketing News*, February 4, p. 4.

Jelen, T. (1992). Political Christianity: A Contextual Analysis. *American Journal of Political Science*. **36**, pp.692–714.

Kellner, P. (2004) Can online polls produce accurate findings? *International Journal of Market Research*, **46**, 2, pp.2-21.

Koch, N. & Emrey, J. (2001). The Internet and Opinion Measurement: Surveying Marginalized Populations. *Social Science Quarterly*. **82:1**. pp. 131-138.

Martzoukou, K. (2005) A review of Web information seeking research: considerations of method and foci of interest. *Information Research*, **10**, 2, paper 215
(<http://www.InformationR.net/ir/10-2/paper215.html>)

Noaman, A. (2006) Pick Me! Pick Me! *ABA Bank Marketing*, **38**, 7, pp. 34-38.

Olson, K. (2006). Survey participation, nonresponse bias, measurement error bias, and total bias. *Public Opinion Quarterly*, **70**, 5, pp.737-758

Reed, J. (1975). Needles in Haystacks: Studying “Rare” populations by secondary

- analysis of national sample surveys. *Public Opinion Quarterly*. **39**, pp.514-522
- Robinson, J. & Godbey, G. (1997). *Time for Life: The Surprising Ways Americans Use Their Time*. University Park, PA: The Pennsylvania State University Press.
- Rothbart, G., Fine, M. & Sudman, S. (1982). On Finding and Interviewing the Needles in the Haystack: The Use of Multiplicity Sampling. *Public Opinion Quarterly*. **46**, pp. 408–21.
- Sharot, T. (2010) Getting back in the frame. *International Journal of Market Research*, **52**, 2, pp.151-153.
- Steeh, C. (1981) Trends in Nonresponse Rates, 1952-1979. *Public Opinion Quarterly* **45**, pp. 40-57.
- Sudman, S. & Blair, E (1999) Sampling in the Twenty-First Century. *Journal of the Academy of Marketing Science*. **27**, 2, pp. 269-278.
- Survey Monkey. (2009) Survey Monkey About Us.
(<http://www.surveymonkey.com/AboutUs.aspx>)
- Taylor, H. (2000) Does internet research work? *International Journal of Market Research*, **42**, 1, pp. 51–63.
- Telang, R., Rajan, U., & Mukhopadhyay, T. (2004) The Market Structure for Internet Search Engines. *Journal of Management Information Systems* **21**, 2, pp. 137-160.
- Thurrow, S. (2006) Debunking SEO Myths.

(<http://www.clickz.com/showPage.html?page=3587096>).

Tourangeau, R. (2006) Survey non-response: a decision-making approach. ESRC Research Methods Festival 2006, St Catherine's College, Oxford.

Wald, K., Owen, D. & Hill, S. (1988). Churches as Political Communities. *American Political Science Review*. **82**, pp. 531–48.

Yost, B. Harding, J., Abbott, C. & Knittle, A. (2005) The Do Not Call Registry: Friend or Foe? The Effect of Do Not Call Lists on Survey Response. American Association for Public Opinion Research Annual Conference.